Particle Physics Data Grid Collaboratory

Data Grids for Science, May 2005

Physics and the Data Grid

The Particle Physics Data Grid (PPDG) provides a pragmatic forum for six experiments and four computer science groups to work together on achieving physics data processing and analysis on production data grid infrastructures.

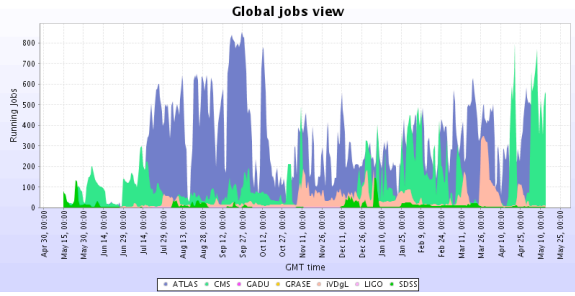
The experiments directly participating in PPDG and other related HENP experiments have successfully transitioned portions of their core data processing to production grids. The physics community is committed to ongoing collaboration to a further vision of a national production grid infrastructure open to contributions by any science.

The close collaboration between computer scientists and the mission-driven experimental physics community has guided new developments of technologies and software tools.

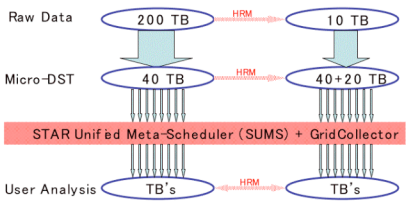
Physics Application Benefits

Particular benefit to physics from deploying data grids comes from collaboration:

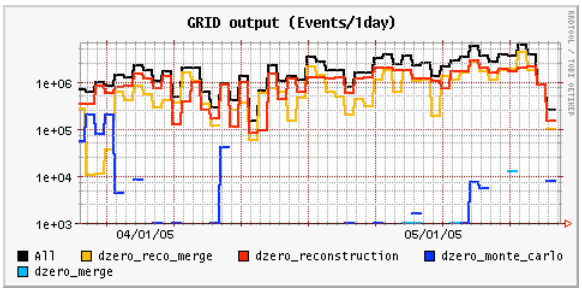
1. Grid3 has been operating as a coherent grid infrastructure with up to 35 sites, 4500 CPUs and 10 applications since November 2003. Grid3 has given the US LHC experiments 30-50% more throughput through opportunistic use of resources they did not own.



2. STAR working with the Storage Resource Management (SRM) and Globus GridFTP for sustained data transfer over the past 2 years:



3. Remote reprocessing of D0 raw data using Condor-G grid scheduler extensions.

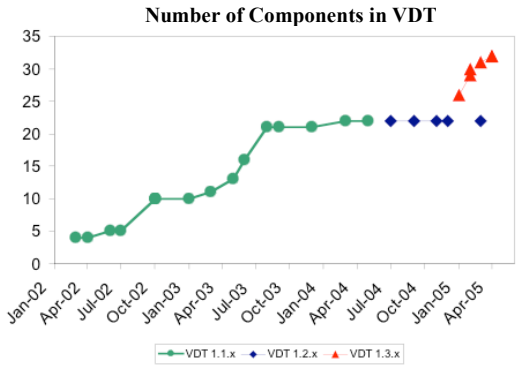


4. Sustained data movement of 100 TBs of BaBar data to remote processing centers driving an extension to the Storage Resource Broker (SRB) to support federated MCAT catalogs.

Computer Science Benefits

The Computer Science Groups have benefited from the collaborations to evolve their technologies in the deployment of services to a broad community. All HENP and Grid3 application communities have adopted the Virtual Data Toolkit (VDT) as the middleware packaging and distribution mechanism. The VDT is delivered by the GriPhyN and iVDGL Grid projects and extended by PPDG, through contributions from the science groups, as well as from the European Data Grid. The VDT relies on the National Middleware Initiative (NMI) build

and test infrastructure, and a coherent and consistent middleware platform is now available.

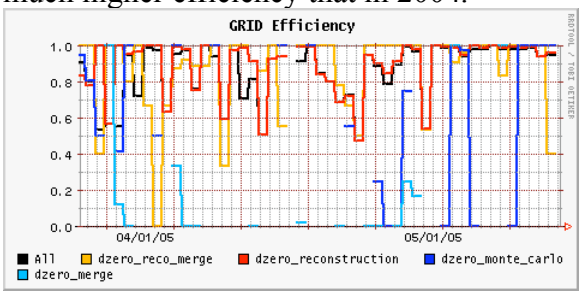


Other examples of PPDG collaborations include: Use of DOE Science Grid certificate-based authentication at all the facilities with hundreds of user certificates and thousands of host certificates issued. Interoperability and compatibility of services and applications with European Grids.

Increasing Robustness

Clearly the success of shared, common Data Grids depends on the efficiency and robustness of the infrastructure. Since the initial capabilities were deployed after the first few years of PPDG there has been a focus on improving the efficiency and throughput.

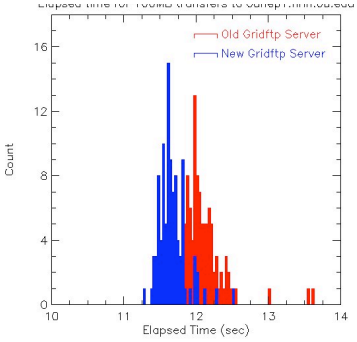
Progress has been made on many fronts in improving the robustness of the Grid infrastructure: D0 reprocessing in 2005 has a much higher efficiency that in 2004.



US ATLAS simulation production job efficiency has been raised from the initial 66%:

ATLAS DC2 Overview of Grids as of 2005-02-24 18:11:30

Grid	submit	pending	running	finished	failed	efficiency
Grid3	36	3	814	**153028**	46943	**77 %**
NorduGrid	29	130	1105	114264	70349	62 %
LCG	60	528	610	145692	242247	38 %
TOTAL	125	661	2529	412984	359539	53 %



New middleware components improve performance. As an example, recent tests of GridFTP on the shared infrastructure show a sustained increase in throughput:

Towards a Production Grid Infrastructure.

As a result of the successes in PPDG and elsewhere, the community has come together to form a consortium of the laboratory-based computer centers, university facilities, application communities and Grid projects serving high energy and nuclear physics to join together in an Open Science Grid (OSG) to deploy a U.S. national persistent Grid infrastructure for science. OSG is being built as a heterogeneous infrastructure relying on common interfaces and service definitions to achieve coherence and interoperation of diverse resources and interoperability between campus, international and other comparable grids. At the same time services in support for multi-user, chaotic data access analysis applications are being added to the common grid infrastructure.

The first version of OSG is now almost ready for deployment with an extended set of services and applications from those on Grid3.